



United States Patent and Trademark Office

UNITED STATES DEPARTMENT OF COMMERCE United States Patent and Trademark Office Address: COMMISSIONER FOR PATENTS P.O. Box 1450 Alexandria, Virginia 22313-1450 www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.	
09/818,607	03/28/2001	Yasuo Okutani	35.G2761	1901	
5514 7	5514 7590 04/16/2004			EXAMINER	
FITZPATRICK CELLA HARPER & SCINTO			WOZNIAK	WOZNIAK, JAMES S	
••	30 ROCKEFELLER PLAZA NEW YORK, NY 10112		ART UNIT	PAPER NUMBER	
			2655	6	
			DATE MAILED: 04/16/2004		

Please find below and/or attached an Office communication concerning this application or proceeding.

	Application No.	Applicant(s)				
	•					
Office Action Summary	09/818,607 Examiner	OKUTANI ET AL. Art Unit				
•	James S. Wozniak	2655				
The MAILING DATE of this communication						
Period for Reply	• •	, , , , , , , , , , , , , , , , , , , ,				
A SHORTENED STATUTORY PERIOD FOR F THE MAILING DATE OF THIS COMMUNICAT - Extensions of time may be available under the provisions of 37 of after SIX (6) MONTHS from the mailing date of this communicat. - If the period for reply specified above is less than thirty (30) days of 18 of 19	CION. CFR 1.136(a). In no event, however, may a re ion. s, a reply within the statutory minimum of thirty period will apply and will expire SIX (6) MONT at statute, cause the application to become ABA	eply be timely filed y (30) days will be considered timely. THS from the mailing date of this communication. ANDONED (35 U.S.C. § 133).				
Status						
1) Responsive to communication(s) filed on	03/03/04.					
3) Since this application is in condition for a						
closed in accordance with the practice under Ex parte Quayle, 1935 C.D. 11, 453 O.G. 213.						
Disposition of Claims						
4)⊠ Claim(s) <u>1-7,9-19 and 21-25</u> is/are pending in the application.						
•	4a) Of the above claim(s) is/are withdrawn from consideration.					
5) Claim(s) is/are allowed.						
6)⊠ Claim(s) <u>1-7, 9-19, & 21-25</u> is/are rejecte	·					
7) Claim(s) is/are objected to.	7) Claim(s) is/are objected to.					
8) Claim(s) are subject to restriction	and/or election requirement.					
Application Papers						
9)☐ The specification is objected to by the Examiner.						
10)⊠ The drawing(s) filed on <u>28 March 2001</u> is/are: a)⊠ accepted or b)□ objected to by the Examiner.						
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).						
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).						
11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.						
Priority under 35 U.S.C. § 119						
12)⊠ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).						
a)⊠ All b)□ Some * c)□ None of:						
1. Certified copies of the priority documents have been received.						
2. Certified copies of the priority documents have been received in Application No						
3. Copies of the certified copies of the	e priority documents have been r	received in this National Stage				
application from the International E	Bureau (PCT Rule 17.2(a)).					
* See the attached detailed Office action for a list of the certified copies not received.						
Attachment(s)						
1) Notice of References Cited (PTO-892) 2) Notice of Draftsperson's Patent Drawing Review (PTO-948) 4) Interview Summary (PTO-413) Paper No(s)/Mail Date						
3) Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) 5) Notice of Informal Patent Application (PTO-152)						
Paper No(s)/Mail Date 6) Other:						

Detailed Action

Response to Amendment

1. In response to the office action from 12/03/03, the applicant has submitted an amendment, filed 3/03/04, requesting that the provisional double patenting over co-pending application no. 09/818,581 be held in abeyance, amending Claims 1, 4-7, 9, 10, 12, 13, 16-19, 21, 22, and 21 without adding new matter, and canceling Claims 8 and 20, while arguing to traverse the art rejection over the amended claims based on the limitation regarding the selection of synthesis units based on modification distortion (Amendment, Page 11). The rejection of dependent claims have been argued only for being dependent upon an independent claim with no additional arguments for limitations allegedly not being taught by the references cited.

Applicant's arguments with respect to claims 1 and 13 have been considered but are moot in view of the new ground(s) of rejection due to the amended claims, based on Kagoshima et al (U.S. Patent: 6,240,384):

Double Patenting

2. The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. See *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed.

Cir. 1993); In re Longi, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); In re Van Ornum, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); In re Vogel, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); and, In re Thorington, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground provided the conflicting application or patent is shown to be commonly owned with this application. See 37 CFR 1.130(b).

Effective January 1, 1994, a registered attorney or agent of record may sign a terminal disclaimer. A terminal disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

3. Claims 1-7, 9-19, and 21-25 are provisionally rejected under the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 1-21 of copending Application No. 09/818,581. Although the conflicting claims are not identical, they are obvious variations of one another because both disclose a speech synthesis system utilizing concatenation and modification distortion in the selection a best instance of a speech unit, and Application No. 09/818,581 employs the use of Nbest processing in the selection process, a difference which, to one of ordinary skill in the art, at the time of invention, would have been obvious since an Nbest processing method would be more time effective in selecting a best instance. Nbest processing allows only the best speech candidates to be examined instead of an entire speech database, thus resulting in a reduction in processing time.

This is a <u>provisional</u> obviousness-type double patenting rejection because the conflicting claims have not in fact been patented.

Art Unit: 2655

Claim Rejections - 35 USC § 102

Page 4

4. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless -

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

5. Claims 1, 2, 4, 12-14, 16, and 24 are rejected under 35 U.S.C. 102(e) as being unpatentable over Kagoshima et al (U.S. Patent: 6,240,384).

With respect to Claims 1 and 13, Kagoshima discloses:

A speech signal processing apparatus and method comprising:

Distortion obtaining means for obtaining a modification distortion between synthesis units before and after modification (distortion calculator for determining a distortion between a synthesis speech segment and a training speech segment, Col. 13, Lines 58-60. Also, the training speech segment is modified with respect to pitch and duration to generate a synthesis speech segment, Col. 8, Lines 62-66).

Selection means for selecting synthesis units based on the modification distortion obtained by said distortion obtaining means (selecting synthesis units that minimize distortion based on a distance comparison between synthesis and training units, Col. 2, Lines 58-62); and

Speech synthesis means for performing speech synthesis based on the synthesis units selected by said selection means (speech synthesizer, Fig. 1, Element 15).

Art Unit: 2655

With respect to Claims 2 and 14, Kagoshima recites:

A speech signal processing apparatus and method, wherein said selection means selects a plurality of synthesis units based on a phoneme series including a plurality of phonemes (selected synthesis units which correspond to phonetic clusters, Col. 3, Lines 9-13).

With respect to Claims 4 and 16, Kagoshima discloses:

A speech signal processing apparatus, wherein the selection means selects the synthesis units so as to minimize the modification distortion (selection of synthesis speech segments that will reduce distortion, Col. 2, Lines 58-62).

With respect to Claims 12 and 24, Kagoshima recites:

Input means and step for inputting text data (input text, Col. 8, Line 10, that would inherently be inputted via a text input means);

Language analysis means and step for performing language analysis of the text data (language processing of an input text, Col. 15, Lines 41-43); and

Prosody-parameter generation means and step for generating predetermined prosody parameters based on a result of analysis of said language analysis means and step (obtaining prosody information from language processing, Col. 15, Lines 41-43).

Wherein said distortion obtaining means obtains the modification distortion between the synthesis units before and after modification based on the predetermined prosody parameters generated by said prosody parameter generation means (distortion calculator for determining a distortion between a synthesis speech segment (training segment with added prosody information) and a training speech segment, Col. 13, Lines 58-60. Also, the training speech

Art Unit: 2655

segment is modified with respect to pitch and duration to generate a synthesis speech segment, Col. 8, Lines 62-66, according to prosody information, Fig. 1, Element 111).

Page 6

Claim Rejections - 35 USC § 103

- 6. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:
 - (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.
- 7. Claims 3, 9, 15, and 21 are rejected under 35 U.S.C. 103(a) as being unpatentable over Kagoshima et al in view of Campbell et al (U.S. Patent: 6,366,883).

With respect to Claims 3 and 15, Kagoshima discloses the speech synthesis apparatus and method that utilizes a modification distortion in selecting a best speech unit for synthesizing speech, as applied to Claims 1 and 13. Kagoshima does not teach synthesis units corresponding to a single phoneme that are selected based upon the distortion associated with that particular frame, however Campbell recites:

A speech signal processing apparatus and method, wherein said distortion obtaining means obtains a distortion which may be generated in each of a plurality of synthesis units corresponding to one phoneme, and wherein said selection means selects one synthesis unit from among the plurality of synthesis units corresponding to the one phoneme (speech unit selection process of an N best group of individual phonemes, Fig. 7).

Kagoshima and Campbell are analogous art because they are from a similar field of endeavor in speech synthesis. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the partitioning of individual speech units as individual phonemes in a best unit selection process as taught by Campbell with the best instance selection method utilizing modification distortion as taught by Kagoshima to create a speech synthesis method capable of producing a high quality speech sequence with minimized distortion on an individual phoneme basis. Therefore, it would have been obvious to combine Campbell with Kagoshima for the benefit of obtaining a speech synthesis system capable of producing high quality synthesized speech by selecting individual phonemes for concatenation based upon distortion, to obtain the invention as specified in Claims 3 and 15.

With respect to **Claims 9 and 21**, Kagoshima teaches the speech synthesis apparatus and method that utilizes a modification distortion, calculated as the distance between a training and synthesis unit, in selecting a best speech unit for synthesizing speech, as applied to Claims 1 and 13. Kagoshima does not specifically suggest calculating modification distortion using a cepstrum distance, however Campbell discloses:

A speech signal processing apparatus and method, wherein said distortion obtaining means calculates the modification distortion using a cepstrum distance (distortion calculation based upon prosodic feature parameters calculated from acoustic characteristics of speech units, namely, cepstral distance, Col. 12, Lines 1-36).

Kagoshima and Campbell are analogous art because they are from a similar field of endeavor in speech synthesis. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the means of calculating distortion through cepstral

distance as taught by Campbell with the speech synthesis apparatus and method that utilizes a modification distortion, calculated as the distance between a training and synthesis unit, in selecting a best speech unit for synthesizing speech as taught by Kagoshima to create a speech synthesis system in which modification distortion is calculated using cepstral distance, since cepstral distance is a specific example of the distance calculation taught by Kagoshima and a good way to describe a speech unit. Therefore, it would have been obvious to combine Campbell with Kagoshima for the benefit of obtaining a speech synthesis system capable of producing a speech unit, better described through cepstral distance thus producing higher quality synthesized speech, to obtain the invention as specified in Claims 9 and 21.

8. Claims 5, 6, 17, and 18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Kagoshima et al in view of Huang et al (U.S. Patent: 5,913,193).

With respect to **Claims 5 and 17**, Kagoshima teaches the speech synthesis apparatus and method that utilizes a modification distortion in selecting a best speech unit for synthesizing speech, as applied to Claims 1 and 13. Kagoshima does not teach obtaining a concatenation distortion, however Huang discloses:

A speech signal processing apparatus and method, wherein the distortion obtaining means obtains distortion generated by concatenating a synthesis unit to another synthesis unit (spectral distortion between adjacent instances, Col. 3, Lines 1-6).

Kagoshima and Huang are analogous art because they are from a similar field of endeavor in speech synthesis. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the method of obtaining a concatenation distortion as

taught by Huang with the speech synthesis apparatus and method that utilizes a modification distortion in selecting a best speech unit for synthesizing speech as taught by Kagoshima to further provide more natural synthesized speech by selecting a best synthesis unit additionally based upon concatenation distortion, thus minimizing distortion due to concatenation and creating smooth transitions between speech units. Therefore, it would have been obvious to combine Huang with Kagoshima for the benefit of obtaining a speech synthesis system capable of producing more natural synthesized speech by selecting best speech instances based upon concatenation and modification distortion, to obtain the invention as specified in Claims 5 and 17.

With respect to Claims 6 and 18, Kagoshima teaches the speech synthesis apparatus and method that utilizes a modification distortion in selecting a best speech unit for synthesizing speech, as applied to Claims 1 and 13. Kagoshima does not teach obtaining a distortion by adding modification and concatenation distortion, however Huang discloses:

A speech signal processing apparatus and method, wherein the distortion obtaining means uses a value obtained by adding the modification distortion between the synthesis units before and after modification and a concatenation distortion (spectral distortion between adjacent instances, Col. 3, Lines 1-6) generated by concatenating a synthesis unit to another synthesis unit (summing the distortions of an instance sequence, Col. 9, Lines 44-47).

Kagoshima and Huang are analogous art because they are from a similar field of endeavor in speech synthesis. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the method of summing distortions including a concatenation distortion as taught by Huang with the speech synthesis apparatus and method that

Page 10

utilizes a modification distortion in selecting a best speech unit for synthesizing speech as taught by Kagoshima to further provide more natural synthesized speech by selecting a best synthesis unit dually based upon concatenation and modification distortion, thus minimizing distortion due to concatenation to create smooth transitions between speech units and modification to ensure natural sounding speech in the instance of a prosody change. Therefore, it would have been obvious to combine Huang with Kagoshima for the benefit of obtaining a speech synthesis system capable of producing more natural synthesized speech by selecting best speech instances based upon concatenation and modification distortion, to obtain the invention as specified in Claims 6 and 18.

9. Claims 7 and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Kagoshima et al in view of Huang et al, and in further view of Campbell et al.

With respect to Claims 7 and 19, Kagoshima in view of Huang teaches the speech synthesis system capable of selecting best speech instances based upon a concatenation and modification distortion sum, as applied to Claims 6 and 18. Kagoshima in view of Huang does not teach calculating a distortion as a weighted sum of modification and concatenation distortion, however Campbell discloses:

A speech signal processing apparatus and method, wherein the distortion obtaining means calculates a weighted sum of the modification distortion between the synthesis units before and after modification and the concatenation distortion generated by concatenating a synthesis unit to another synthesis unit (selecting a speech unit based upon weighted coefficient vectors, Col. 2, Lines 37-38).

Page 11

Kagoshima, Huang, and Campbell are analogous art because they are from a similar field of endeavor in speech synthesis. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the method of selecting a speech unit based upon a weighted coefficient vector as taught by with the speech synthesis system capable of selecting best speech instances based upon a concatenation and modification distortion sum as taught by Kagoshima in view of Huang to provide a means of minimizing concatenation cost expressed through a weighting function and thus providing higher quality and audible synthesized speech. Therefore, it would have been obvious to combine Campbell with Kagoshima in view of Huang for the benefit of obtaining a speech synthesis system capable of producing high quality and audible synthesized speech by weighting a distortion sum, to obtain the invention as specified in Claims 7 and 19.

10. Claims 10 and 22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Kagoshima et al in view of Hon et al (U.S. Patent: 6,490,563).

With respect to Claims 10 and 22, Kagoshima teaches the speech synthesis apparatus and method that utilizes a modification distortion in selecting a best speech unit for synthesizing speech, as applied to Claims 1 and 13. Kagoshima does not teach the use of a table to determine a distortion, however Hon discloses:

A speech signal processing apparatus and method, wherein the distortion obtaining means includes a table storing distortions, and determines the modification distortion by referring to the table (use of a unit inventory that contains speech instances and a decision tree that denotes the best speech instances with regard to a joint distortion function consisting of a concatenation and

Application/Control Number: 09/818,607 Page 12

Art Unit: 2655

prosody distortion, both of which may be stored in memory, Col. 6, Line 58- Col 7, Line 5).

Kagoshima and Hon are analogous art because they are from a similar field of endeavor in speech synthesis. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of an inventory and decision tree denoting speech instances with respect to concatenation and prosodic distortion in selecting a best speech instance as taught by Hon with the speech synthesis apparatus and method that utilizes a modification distortion in selecting a best speech unit for synthesizing speech as taught by Kagoshima to create a means of saving distortion parameters for instances where similar text inputs exist- a stored distortion in an inventory and best instance saved in a decision tree could be looked up easily and be used for selecting the best speech instance, thus improving processing speed without degrading speech quality. It would also have been obvious to one of ordinary skill in the art, at the time of invention, to implement the inventory in a lookup table format, as is well known in the art, so that the speech unit with the least distortion could be selected. Therefore, it would have been obvious to combine Hon with Kagoshima for the benefit of obtaining a speech synthesis system with improved processing speed through the use of a speech unit with an associated distortion lookup table for selecting the best speech instance, to obtain the invention as specified in Claims 10 and 22.

11. Claims 11 and 23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Kagoshima et al in view of Huang et al, and in further view of Hon et al.

With respect to Claims 11 and 23, Kagoshima in view of Huang teaches the speech synthesis system capable of selecting best speech instances based upon a concatenation and

Art Unit: 2655

modification distortion, as applied to Claims 5 and 17. Kagoshima in view of Huang do not teach the use of a table to determine a distortion, however Hon discloses:

Page 13

A speech signal processing apparatus and method, wherein the distortion obtaining means includes a table storing distortions, and determines the modification distortion by referring to the table (use of a unit inventory that contains speech instances and a decision tree that denotes the best speech instances with regard to a joint distortion function consisting of a concatenation and prosody distortion, both of which may be stored in memory, Col. 6, Line 58- Col 7, Line 5).

Kagoshima, Huang, and Hon are analogous art because they are from a similar field of endeavor in speech synthesis. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of an inventory and decision tree denoting speech instances with respect to concatenation and prosodic distortion in selecting a best speech instance as taught by Hon with the speech synthesis system capable of selecting best speech instances based upon a concatenation and modification distortion as taught by Kagoshima in view of Huang to create a means of saving distortion parameters for instances where similar text inputs exist- a stored distortion in an inventory and best instance saved in a decision tree could be looked up easily and be used for selecting the best speech instance, thus improving processing speed without degrading speech quality. It would also have been obvious to one of ordinary skill in the art, at the time of invention, to implement the inventory in a lookup table format, as is well known in the art, so that the speech unit with the least distortion could be selected. Therefore, it would have been obvious to combine Hon with Kagoshima in view of Huang for the benefit of obtaining a speech synthesis system with improved processing speed through the use of a speech

unit with an associated distortion lookup table for selecting the best speech instance, to obtain the invention as specified in Claims 11 and 23.

12. Claim 25 is rejected under 35 U.S.C. 103(a) as being unpatentable over Kagoshima et al in view of Huang et al and Campbell et al, and in further view of Hon et al.

With respect to Claim 25, Kagoshima in view of Huang and in further view of Campbell teaches the speech synthesis system capable of selecting best speech instances based upon a concatenation and modification distortion, as applied to Claims 13-24. Kagoshima in view of Huang and in further view of Campbell does not specifically suggest method implementation using a storage medium, however, Hon discloses:

A storage medium, capable of being read by a computer, storing a program for executing a speech signal processing method (computer readable storage medium containing computer instructions for implementing speech synthesis, Col. 4, Lines 36-39).

Kagoshima, Huang, Campbell, and Hon are analogous art because they are from a similar field of endeavor in speech synthesis. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of a computer readable medium for implementing a speech synthesis method as taught by Hon with the speech synthesis system capable of selecting best speech instances based upon a concatenation and modification distortion as taught by Kagoshima in view of Huang because it would have been obvious to one of ordinary skill in the art, at the time of invention, to store the speech processing method taught by Kagoshima in view of Huang on a computer readable medium to increase method compatibility and usability by providing a means for method use with multiple computer

systems. Therefore, it would have been obvious to combine Kagoshima, Huang, Campbell, and Hon for the benefit of obtaining a speech synthesis method executable using a computer, to obtain the invention as specified in Claim 25.

Conclusion

13. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

14. Any inquiry concerning this communication or earlier communications from the examiner should be directed to James S. Wozniak whose telephone number is (703) 305-8669 and email is James. Wozniak@uspto.gov. The examiner can normally be reached on Mondays-Fridays, 8:30-4:30.

Page 16

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Ivars Smits can be reached at (703) 306-3011. The fax/phone number for the Technology Center 2600 where this application is assigned is (703) 872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the technology center receptionist whose telephone number is (703) 306-0377.

James S. Wozniak 4/14/2004

DORIS H. TO SUPERVISORY PATENT EXAMINER TECHNOLOGY CENTER 2600